Statistical Methods in Spectral Estimation

Rainer Dahlhaus

1. Introduction

Suppose a certain variable X_t is measured at discrete equally spaced time points $t = 1, \dots, T$ and we want to make assertions on the energy distribution of the frequencies in a harmonic analysis. This energy distribution may for example be used to code the data if X_t is a speech signal or to make a prediction if the X_t are economic data.

Instead of using a deterministic approach applied scientists usually use a stochastic approach to model the data and to estimate the energy distribution (e.g. in electrical engineering, geophysics, economics or neurophysiology). One reason is that in a stochastic setup certain fluctuations of the Fourier-transform of the data can be interpreted more naturally (cp. section 2).

In this paper we discuss the estimation of the energy distribution in a stochastic model and, more generally, the use of the energy distribution for other purposes, such as the detection of hidden frequencies, the estimation of parameters in parametric models and the prediction of the series.

The stochastic framework of this paper is the theory of stationary processes. A stationary process is a stochastic process whose finite dimensional distributions are time invariant (c.f. Brillinger, 1981, section 2.4). In particular the mean and the correlation of neighbouring values of a stationary process remain constant over time.

An important feature of stationary processes is the following spectral representation (cf. Brockwell and Davies, 1991, Theorem 4.8.2).

(1.1) Spectral representation

Let X_t, $t \in \mathbf{Z}$ be a stationary sequence with mean zero. Then there exists a stochastic process $\xi(\lambda)$, $\lambda \in [-\pi,\pi]$ with $\xi(\lambda) = \xi(-\lambda)$ and orthogonal increaments such that

$$X_t = \int_{-\pi}^{\pi} \exp(i\lambda t)\, d\xi(\lambda). \tag{1.2}$$

The spectral distribution function $F(\lambda) := \mathbf{E}|\xi(\lambda) - \xi(-\pi)|^2$ is right-continuous and non-decreasing.

---

[1] This paper was presented in the final colloquium of the Sonderforschungsbereich 123 “Stochastische Mathematische Modelle” in Heidelberg, December 12, 1992.

The derivative of F (with respect to the Lebesgue measure) is called spectral density and denoted by $f(\lambda)$ (provided it exists). If $f(\lambda) > 0$ then we also have a representation

$$X_t = \int_{-\pi}^{\pi} \exp(i\lambda t)\, A(\lambda)\, d\xi(\lambda) \tag{1.3}$$

with $f(\lambda) = |A(\lambda)|^2$ and a process $\xi(\lambda)$ with orthogonal increaments, which, in addition, fulfills $\mathbf{E}|\xi(\lambda) - \xi(\mu)|^2 = \lambda - \mu \quad (\lambda \geq \mu)$.

Thus, if we model the observations $X_1, \ldots, X_T$ by a stationary process, the estimation of the energy of the harmonic components means estimation of $f(\lambda) = |A(\lambda)|^2$. Before dealing with this problem we give an important example of a stationary process.

(1.4) Autoregressive processes (AR-processes)

Consider the linear recursion (frequently termed as linear difference equation).

$$\sum_{j=0}^{p} a_j X_{t-j} = \varepsilon_t \quad (a_0 = 1) \tag{1.5}$$

where ε_t are independently and identically distributed random variables with mean zero and variance σ^2. The corresponding homogeneous equation

$$\sum_{j=0}^{p} a_j X_{t-j} = 0$$

has a solution of the form

$$X_t = \sum_{j=1}^{l} u_j^t \sum_{i=1}^{m_j - 1} b_{ij}\, t^i$$

where

$$\sum_{j=0}^{p} a_j z^j = \prod_{j=1}^{l} (1 - u_j z)^{m_j} \qquad u_j \in \mathbf{C} \ .$$

If $u_j = \theta_j e^{-i\mu_j}$ with $\theta_j \in \mathbf{R}_+$ then (1.5) has a stationary causal solution if $\theta_j < 1$. In this case u_j^t is a damped oscillation with frequency μ_j. The random shocks ε_t prevent the series X_t from tending to zero. The spectral density of this process is

(1.6) $$f(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{p} a_j \, e^{i\lambda j} \right|^{-2} = \frac{\sigma^2}{2\pi} \prod_{j=1}^{l} \left| 1 - \theta_j \, e^{i(\lambda - \mu_j)} \right|^{-2m_j} .$$

For θ_j close to 1 $f(\lambda)$ has a peak at frequency μ_j. Figure 1 shows a realisation $X_1, \ldots, X_T$ of length T = 128 of (1.5) where the ε_t are (pseudo-) Gaussian variables with mean 0 and variance 1, p = 4

$$u_1 = 0.9 \, e^{-i\,0.12\pi} , \qquad u_2 = \overline{u_1}$$
$$u_3 = 0.9 \, e^{-i\,0.4\pi} , \qquad u_4 = \overline{u_3} .$$

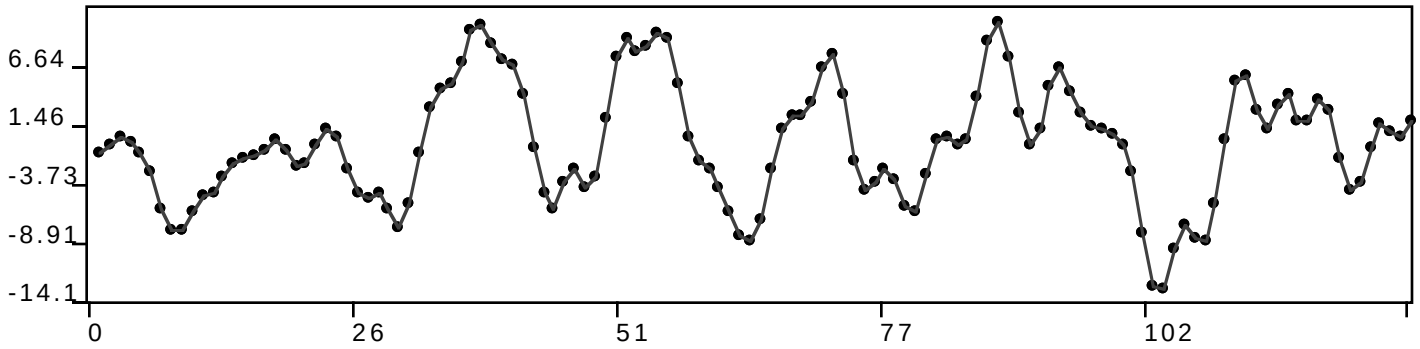

Fig. 1. T=128 realisations of an autoregressive process of order 4

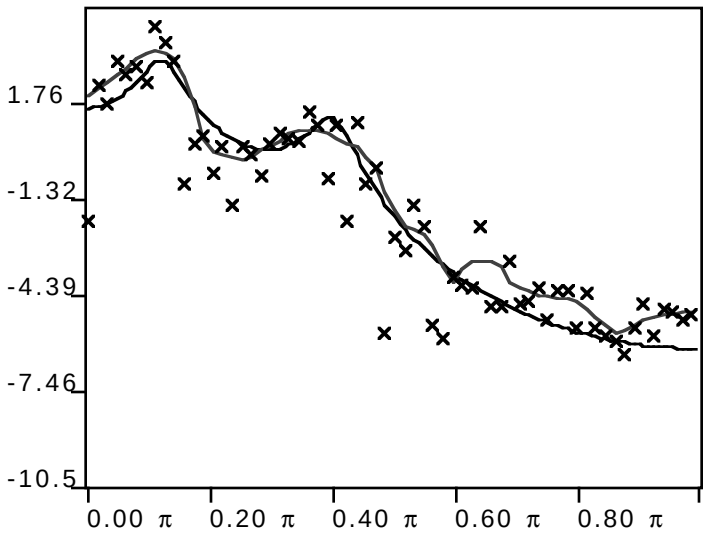

Fig. 2. True spectrum, periodogram, kernel estimate

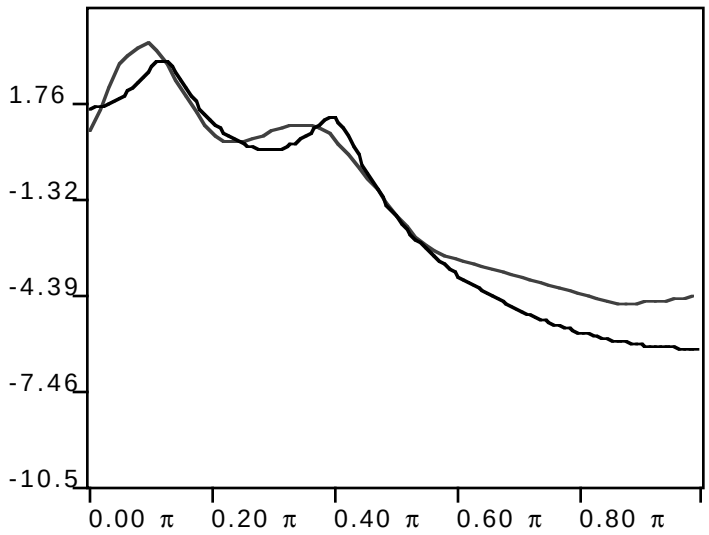

Fig. 3. True spectrum with adaptive kernel estimate

The data clearly show a periodic behaviour at the two frequencies . The dark line in Figure 2 is the spectral density of this process (all spectra are plotted on a logarithmic scale). In practice this spectral density is unknown and has to be estimated from the data set $X_1, \dots, X_T$.

Autoregressive processes are for example frequently used for coding segments of speech signals. In the engineering literature this method is termed LPC-method (linear predictive coding).

In this paper we discuss nonparametric and parametric estimates of the spectral density. Furthermore, we present a bootstrap procedure in the frequency domain and a method for detecting hidden periodicities in the data.

We focus our presentation on the work done in the Teilprojekt B2 of the SFB 123 from 1988 to 1992. Further techniques in spectral estimation may be found e.g. in the monograph of Brillinger (1981).

2. Nonparametric estimation of the spectral density

Equation (1.3) suggests to take the squared inverse Fourier-transform of $X_1, \dots, X_T$ as an estimate of $f(\lambda)$. This estimate is the periodogram

$$I_T(\lambda) := \frac{1}{2\pi T} \left| \sum_{t=1}^{T} X_t \exp(-\lambda t) \right|^2 .$$

The asterisks in Figure 2 are the periodogram ordinates for the process from Figure 1 at frequencies $\lambda_j = \frac{2\pi j}{T}$, $j = 1, \dots, T/2$. The values fluctuate around the true (unknown) spectral density. This is a bit surprising since the original process X_t seems to be quite smooth. In our stochastic approach this behaviour can easily be explained by the distributional properties of the periodogram. Under fairly general conditions we have for $\lambda, \mu \in (0, \pi)$ with $\lambda \neq \mu$

$$\mathbf{E}\, I_T(\lambda) = f(\lambda) + O(T^{-1})$$

(2.1)
$$\operatorname{var} I_T(\lambda) = f(\lambda)^2 + O(T^{-1})$$

$$\operatorname{cov}(I_T(\lambda), I_T(\mu)) = O(T^{-1})$$

(cf. Brillinger, 1981, chapter 5). Moreover, $I_T(\lambda_j)_{j=1,\dots,k}$ are asymptotically independent

$f(\lambda_j)\ \chi^2_2/2$ variates. These results suggest to plot log $I_T(\lambda)$ instead of $I_T(\lambda)$ (as in Figure 2) since the log transformation makes the variance of the estimate (asymptotically) independent of $f(\lambda)$.

In order to obtain a reasonable estimate of $f(\lambda)$ we may e.g. smooth the periodogram ordinates by a weighted moving average (kernel estimate), i.e. we take

$$f_T(\lambda) = \frac{2\pi}{T} \sum_j W_T(\lambda - \frac{2\pi j}{T})\, I_T(\frac{2\pi j}{T}) \tag{2.2}$$

with $W_T(\alpha) := \frac{1}{b_T} W(\frac{\alpha}{b_T})$ where $\int W(\alpha)d\alpha = 1$. b_T is called the bandwidth. $f_T(\lambda)$ is the dotted line in Figure 2 (where W is the Bartlett-Priestley kernel and $b_T = 0.05$ – cp. Priestley, 1981, section 7.5).

Since $I_T(\lambda_j) / f(\lambda_j)$ are asymptotically iid (for finitely many fixed λ_j) we take as a measure for the goodness of fit of a spectral estimate the integrated relative mean squared error

$$\mathrm{IMSE}(f_T) := \int_{-\pi}^{\pi} \mathbf{E}\left(\frac{f_T(\lambda)}{f(\lambda)} - 1\right)^2 d\lambda\ .$$

Using a second order expansion for the bias (cf. Priestley, 1981, Section 7.5) leads to

$$b_T = T^{-1/5}\left[\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{f''(\alpha)}{f(\alpha)}\right)^2 d\alpha\right]^{-1/5}\left(\frac{\int W(y)^2 dy}{\int y^2 W(y) dy}\right)^{1/5}$$

as the optimal bandwidth, and to the Bartlett-Priestley kernel

$$W(y) = \frac{3}{4\pi}\left(1 - \left(\frac{y}{\pi}\right)^2\right), \quad y \in [-\pi, \pi]$$

as the optimal kernel. Unfortunately, this bandwidth depends on the unknown f. Therefore, the question for a data driven bandwidth selection arises.

The presence of sharp peaks and flat regions at the same time causes additional problems. With a small bandwidth $f_T(\lambda)$ is a reasonable estimate at the peaks but too irregular at the flat regions. A larger bandwidth leads to a reasonable estimate at the flat regions but the peaks become to broad . Therefore, a local bandwidth seems preferable.

The theoretically optimal local bandwidth is given by the same expression as above where $(f''(\alpha) / f(\alpha))^2$ is not integrated over α. Again the bandwidth depends on the unknown f and a data driven local bandwidth selection is even more complicated.

The problem is similar to the problem of bandwidth selection in nonparametric regression. The latter problem has been studied in detail in the Teilprojekt B1. Gasser, Kneip and Kohler (1991) have given an adaptive global bandwidth selection procedure and Brockmann (1992) an adaptive local bandwidth selection procedure for nonparametric regression.

Since the variance of the periodogram depends on the unknown spectral density these procedures require some modifications. Figure 3 shows a kernel estimate with the modified local bandwidth selection of Brockmann (dotted line). This estimator clearly has a better behaviour, at the peaks and at the flat regions.

On the other hand the mathematical smoothness assumptions needed for the above optimality properties of kernel estimates can hardly be justified for spectral densities where typically sharp peaks occur. It seems to be better to use the prior knowledge of the possible form of the peaks (sharp peaks) in a semiparametric approach where the peaks are modelled parametrically (e.g. as peaks of an AR-process with coloured noise) and the remainder is smoothed nonparametrically.

If outlier are present in the periodogram (due to periodic components in the data) a robust smoothing technique may be preferred. M-estimates in the frequency domain have been studied by v. Sachs (1990).

We now study the same estimates if the peaks in the spectrum are stronger. We therefore generate an AR-process with p = 10 and double roots

$$u_1 = u_2 = 0.9e^{-i0.12\pi}, \qquad u_3 = u_4 = \overline{u_1}$$
$$u_5 = u_6 = 0.9e^{-i0.4\pi}, \qquad u_7 = u_8 = \overline{u_5}$$

and additional single root

$$u_9 = 0.9e^{-i0.8\pi}, \qquad u_{10} = \overline{u_9}\ .$$

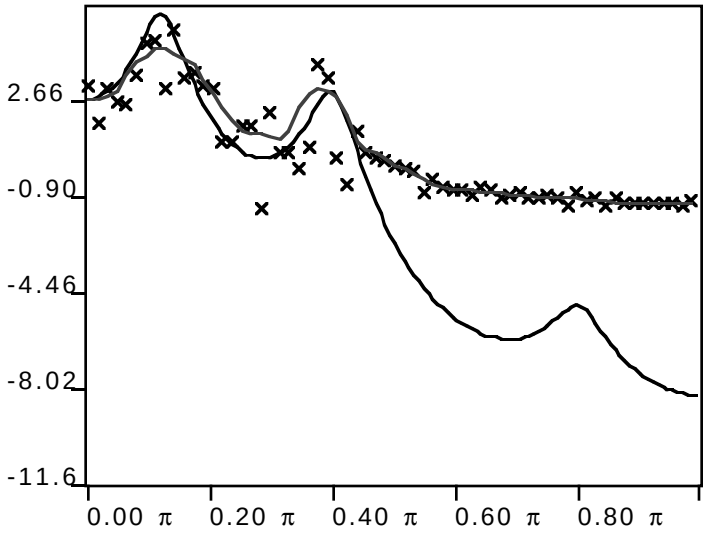

Fig. 4. True spectrum, periodogram, kernel estimate without data-taper

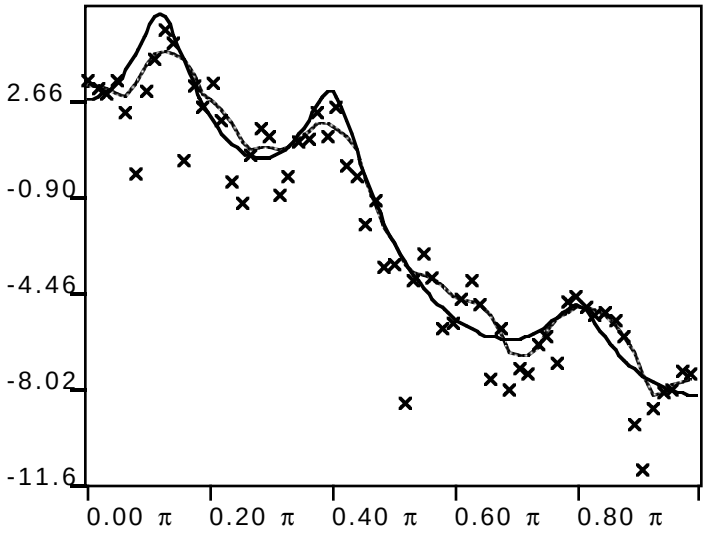

Fig. 5. True spectrum, periodogram, kernel estimate with data-taper

The true spectral density, the periodogram and the adaptive kernel estimate are plotted in Figure 4. We see a strong bias of the periodogram and of the resulting kernel estimate. The true

spectrum is overestimated. In particular the smaller peak at $\lambda = 0.8\pi$ is hidden. This effect is called leakage effect.

A heuristic explanation is the following. Straightforward calculation gives for the expectation of the periodogram

$$\mathbf{E}\, I_T(\lambda) = \int_{-\pi}^{\pi} f(\lambda + \alpha)\, K_T(\alpha) d\alpha \tag{2.3}$$

where

$$K_T(\alpha) = \frac{1}{2\pi T} \frac{\sin^2 (T\alpha)/2}{\sin^2 \alpha/2}$$

is the Fejér kernel, i.e. the expectation of $I_T(\lambda)$ is the convolution of the unknown f with K_T. Since $K_T(\alpha)$ has side maxima at frequencies $\alpha = \frac{(2k + 1)\pi}{T}$ (k = 1,2, …) the energy of f at a peak leaks in the convolution to other frequencies which explains the behaviour of the estimates in Figure 4.

Tukey (1967) suggested the use of data-tapers to improve the behaviour of these estimates. Prior to the Fourier transformation the data are weighted down at the ends of the observation domain, i.e. instead of $I_T(\lambda)$ we use

$$I_T^{(h)}(\lambda) = \frac{1}{2\pi H_{2,T}(0)} \left| \sum_{t=1}^{T} h\left(\frac{t-0.5}{T}\right) X_t \exp(-i\lambda t)\right|^2$$

where

$$H_{k,T}(\lambda) = \sum_{t=1}^{T} h\left(\frac{t-0.5}{T}\right)^k \exp(-\lambda t)$$

and h(x) is a smooth nonnegative function that has its maximum at x = 1/2 and decays smoothly to zero as x tends to 0 or 1. $f_T^{(h)}(\lambda)$is defined as in (2.2) with $I_T^{(h)}$ instead of I_T .

A frequently used taper is the Tukey-Hanning taper where the proportion ρ is tapered with a cosine bell:

$$h_\rho(x) = \begin{cases} 1/2[1-\cos(2\pi\frac{x}{\rho})], & x\in[0,\rho/2] \\ 1, & x\in[\rho/2\ ,\ 1/2] \\ h_\rho(1-x), & x\in\ (1/2,1]\ . \end{cases}$$

If the mean is unknown we take the same expression for $I_T^{(h)}$ as above with $X_t - \overline{X}$ instead of X_t (we remark that tapering the data prior to the calculation of $\overline{X}$ is disadvantageous).

The effect of using a data taper is quite dramatic if the true spectrum contains strong peaks. Figure 5 shows the same estimates as in Figure 4, but with the above Tukey-Hanning taper where ρ = 1.0.

However, it turned out to be very difficult to describe the leakage effect and the benefits of data-tapers theoretically. (2.3) still holds where now

$$K_T(\alpha) = \{2\pi\ H_{2,T}(0)\}^{-1}\ |H_{1,T}(\alpha)|^2\ .$$

A plot of $K_T(\alpha)$ for different data taper shows that it has much smaller side maxima than the Fejér kernel. In fact, one can prove that

$$\mathbf{E}\ I_T^{(h)}(\lambda) = f(\lambda) + O(T^{-2})$$

for a sufficiently smooth data taper while the other properties of (2.1) remain the same (this result only holds if the mean is known). However, the bias of the resulting kernel estimate $f_T^{(h)}$ stays the same while the variance of $f_T^{(h)}$ is increased by a constant factor (cf. Brillinger, 1981, Theorem 5.6.4 and Corollary 5.8.2) which leads to an increase of IMSE(f_T): The leakage effect and the improvement by tapering are small sample effects that disappear in the asymptotic theory.

To overcome these problems we have suggested a refined criterion to judge the quality of spectral estimates (Dahlhaus, 1990). Instead of the IMSE(f_T) we have studied

$$\sup_{X_T} \text{IMSE}(f_T)$$

for several estimates where X_T is a (with T) increasing class of stochastic processes. By using an increasing class one requires that the estimate behaves uniformly good over an increasing number of stochastic processes when the sample size increases. By using such a criterion one can avoid that certain small sample effects such as the leakage effect disappear asymptotically. The definition of the class X_T can be found in Dahlhaus (1990, Section 2). It contains processes with spectral densities whose peaks and troughs increase with T, for example AR-process whose roots u_j tend to the unit circle.

By using this approach it is possible to describe several small sample effects, in particular, to describe the resolution properties of the estimates. The following results hold with the uniform criterion UMSE(f_T): = $\sup_{X_T}$ IMSE(f_T) (cf. Dahlhaus, 1990).

(i) If a smooth data taper and a special local bandwidth is used, we obtain UMSE($f_T^{(h)}$) = O($T^{-4/5}$) (Theorem 3.6). This is the optimal rate even for the IMSE.

(ii) If a global bandwidth is used, we obtain UMSE($f_T^{(h)}$) ≥ $CT^{-2/5}$ (Theorem 4.1).

(iii) If no data taper is used, we obtain UMSE(f_T) ≥ C (Theorem 4.3). This explains theoretically the leakage effect and in addition a similar effect called trough effect (cp. section 4.3).

(iv) It follows (Section 4.4) that also the variance may decrease with a data taper which is contrary to widespread conjectures.

(v) Similar results may be proved for other estimates such as segment estimates (Section 5).

Thus, the advantages of data tapers could also be established theoretically. An important problem is the choice of the taper. No rigorous results exist for this problem. It is obvious, that the choice depends on the true (unknown) spectral density, in particular on the relation of the peaks and troughs to each other.

A heuristic advice that works fairly well in practice is the following resampling procedure. We start by fitting a parametric ARMA-model which catches the important features of peaks and troughs. It is important that for the estimate in this model a high resolution method is used (e.g. a maximum likelihood approach or a minimum distance method with a strong data taper - cp. section 3). Processes of the same length are afterwards generated from the fitted model and the optimal data taper (together with the optimal bandwidth) can be determined in a simulation study.

We remark that there exist other spectral estimates that have similar resolution properties as the tapered periodogram . Those statistics are usually non linear / non quadratic and therefore very difficult to investigate theoretically. An example is the Capon estimate. Ioannidis (1993) has derived the asymptotic properties of several Capon type estimates.

3. Parametric models

An alternative to nonparametric spectral density estimation is the fitting of a parametric spectral density. As an example we now discuss the fitting of an AR-model as in (1.1). In this case the coefficients $a_1, \dots, a_p$ and σ^2 have to be estimated. (1.6) with the estimated parameters then gives the spectral estimate.

One possible approach is to minimize a distance between the theoretical parametric density and the periodogram with respect to the parameters. A natural distance function comes from considering the asymptotic Kullback-Leibler information divergence. It is possible to prove (cf. Parzen, 1983) that the asymptotic information divergence of a process with true spectral density f and a Gaussian model with spectral density f_θ is

$$L(\theta) := \frac{1}{4\pi} \int_{-\pi}^{\pi} \{\log f_\theta(\lambda) + \frac{f(\lambda)}{f_\theta(\lambda)}\} \, d\lambda + \text{constant} . \tag{3.1}$$

Since f is unknown we use the periodogram instead and obtain as an empirical distance

$$(3.2) \qquad L_T(\theta) := \frac{1}{4\pi} \int_{-\pi}^{\pi} \{\log f_\theta(\lambda) + \frac{I_T(\lambda)}{f_\theta(\lambda)} \} \, d\lambda$$

which is the so called Whittle function (Whittle, 1953). Minimizing $L_T(\theta)$ with respect to θ gives the Whittle estimate $\hat{\theta}_T$. $\hat{\theta}_T$ converges to θ_0 - the value which minimizes $L(\theta)$, i.e. the best theoretical fit.

For AR processes f_θ is given by (1.6). In this case $\hat{\theta}_T$ is identical to the solution of the classical Yule-Walker equations (cf. Brockwell and Davies, 1991). It has been known for a long time that Yule-Walker estimates are not very good - in particular the resolution properties are quite bad.

From our discussion it is obvious that an improvement of the nonparametric periodogram in (3.2) will also lead to an improvement of the resulting parameter estimate $\hat{\theta}_T$. Since in (3.2) the periodogram is integrated the use of a smoothed version (kernel estimate) does not lead to any improvement at all. This can be seen in simulations and from theoretical consideration. However, the use of data tapers can have a dramatic effect.

In Figure 6 parametric estimates are plotted for the same series as in Figure 4. The dark line again is the true spectral density (unknown in practice). The light dotted line is the classical estimate $\hat{\theta}_T$ without data-taper while the dark dotted line shows the estimate with a data taper. Again we see the dramatic improvement due to tapering. Furthermore, the special form of the parametric model leads to sharp peaks - no bandwidth is needed. The data-taper may be selected as suggested in Section 2.

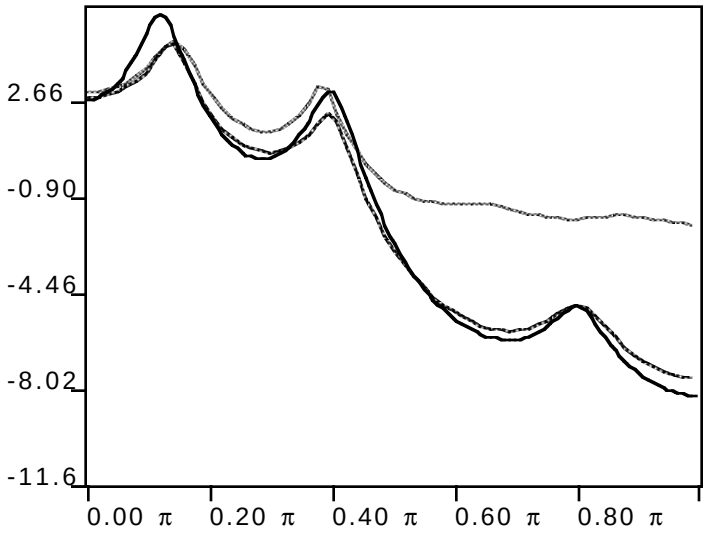

Fig. 6. True spectrum, parametric estimate with and without data-taper

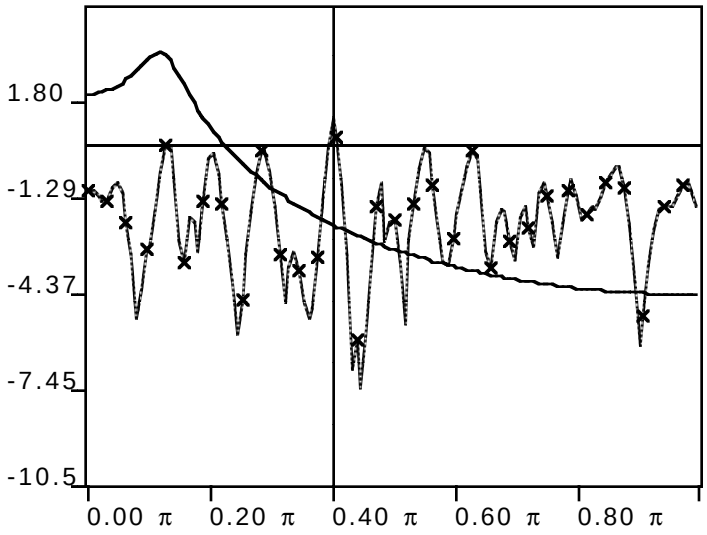

Fig. 7. True spectrum and detection of hidden periodic components

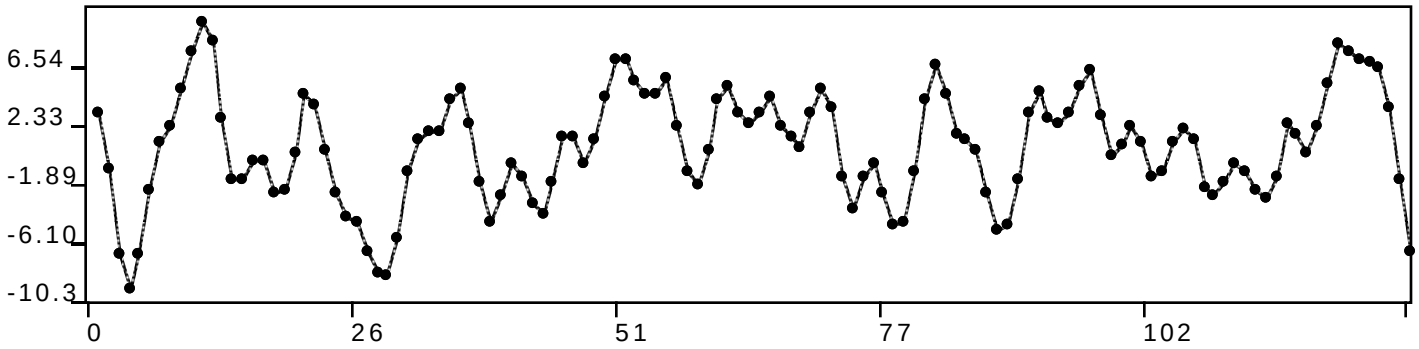

Fig. 8. T=128 realisations of an autoregressive process of order 2 with additional periodic component

Although the nontapered estimate (Yule Walker estimate) has been known for a long time, the knowledge on the benefits of data taper is quite new. This is due to the fact that Yule Walker estimates usually are defined via moment equations and not by the above minimum distance equation. Tapered Yule-Walker estimates were introduced by Dahlhaus (1984). As in the nonparametric case it is very difficult to prove the advantages of data tapers for parametric estimates theoretically. This was done in Dahlhaus (1988) by a similar approach as for the nonparametric case.

We now show by a simulation example how these considerations may be used to improve the prediction of a time series. Figure 9 shows an AR(6) series similar to the one from above. The true spectrum with the nontapered and the tapered estimate are plotted in Figure 10. The series contains two periodicities at frequencies which corresponds to a period of 7 and 30 days which is typical in economic data. Due to leakage the nontapered estimate can only resolve the peak at the frequency that corresponds to the seven days period while the other is hidden.

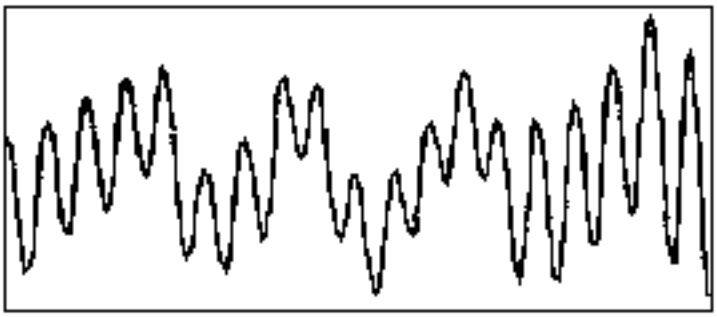

Fig. 9. 128 realisations of an autoregressive process of order 6

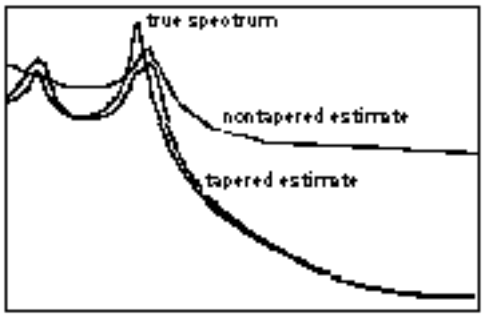

Fig. 10. True spectrum, spectral estimate

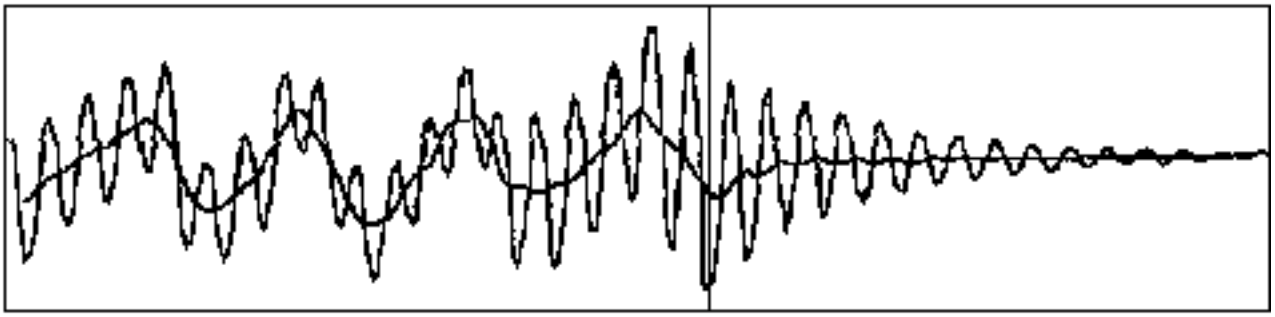

Fig. 11. Autoregressive process with 100 predictions based on a nontapered Yule-Walker estimate

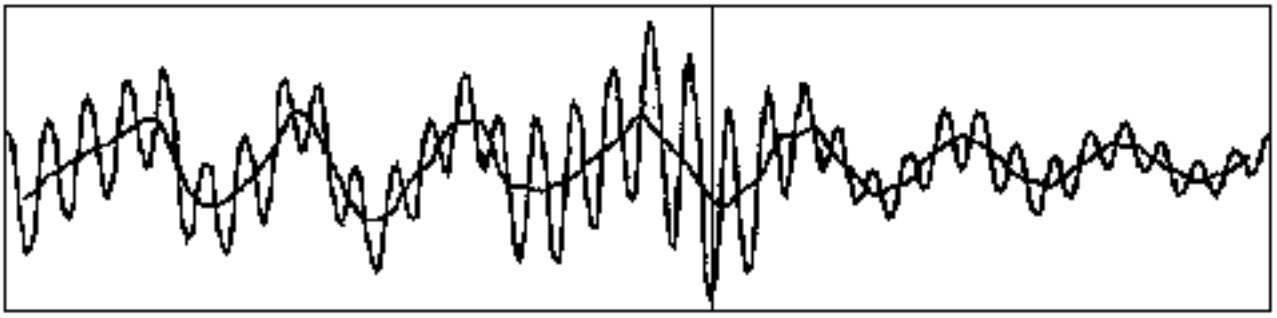

Fig. 12. Autoregressive process with 100 predictions based on a tapered Yule-Walker estimate

The optimal prediction of X_{T+1} based on $X_1, \dots, X_T$ is for an AR-process given by the equation (1.5) with $\varepsilon_t = 0$, namely by

$$\hat{X}_{T+1} = -\sum_{j=1}^{p} a_j X_{T+1-j}$$

(cp. Brockwell and Davies, 1991 , chapter 5.5). The optimal prediction of X_{T+n} is given by the same equation where now the unknown X_t are replaced by their (previous) predictors. To calculate the predictions we now replace the unknown a_j in the prediction equation by their estimates.

Figure 11 shows the series with the predictions if the nontapered Yule-Walker estimate is used. To show the quality of the prediction we have added a moving average over 7 consecutive values. Due to the bad resolution properties of the estimate only the 7-days cycle is predicted while the 30-days cycle is missing in the prediction. If we use the tapered parameter estimate (Figure 12) also the 30-days cycle is predicted sufficiently. (For reasons of clarity we have omitted the confidence bounds of the prediction - in a practical analysis these bounds should be calculated and plotted).

We finally remark that there exist also other parametric estimates that have good resolution properties. Examples are the exact (Gaussian) maximum likelihood estimate and the popular Burg algorithm (Burg, 1967). Hainz (1992) has recently calculated the asymptotic distribution and the asymptotic bias of the Burg estimate. We remark that these estimates can not be improved by using data tapers.

4. Estimation- and detection procedures based on the periodogram

The distributional properties of the periodogram may be used in several directions for statistical problems.

One application is the detection of periodic components in the data. In this case the observations are for example of the form

$$X_t = A \cos(\omega t + \phi) + Y_t \tag{4.1}$$

where Y_t is a stationary process with spectral density $f(\lambda)$. The periodogram of the series X_t has an outlier at frequency ω (in fact also neighboring values are influenced). At frequencies different from ω the asymptotic distribution of $I_T(\lambda) / f(\lambda)$ is $\chi^2_2 /2$. One may use this property to conclude to a periodic component if the value of $I_T(\lambda)/f(\lambda)$ lies e.g. outside the 95% confidence interval of the $\chi^2_2/2$ distribution. The problem is that $f(\lambda)$ is unknown and has to be estimated by some estimate $f_T(\lambda)$. If one uses the kernel estimate $f_T(\lambda)$ from section 2 the whole procedure does not work since $f_T(\lambda)$ itself is corrupted by the outlier. However, if one uses the robust estimate of v. Sachs (1991) the whole procedure becomes a powerful tool for detecting hidden periodicities.

Figure 8 shows a series of the form (4.1) with $\omega = 0.4\pi$ and an AR(2)-process Y_t with roots $u = 0.9^{-i0.12\pi}$ and $\bar{u}$. In Figure 7 $I_T(\lambda) / f_T(\lambda)$ is plotted with the robust estimate $f_T(\lambda)$ of v. Sachs and a threshold which is the 95% confidence bound of the supremum of T/2 iid $\chi^2_2/2$ random variables. At $\lambda = 0.4\pi$ $I_T(\lambda) /f_T(\lambda)$ exceeds the threshold and the periodic component is detected. In the plot of the series the periodic behaviour at $\lambda = 0.12\pi$ (resulting from the root of the AR-part) seems to be even stronger. However, the procedure finds correctly the true periodic component. The same method with an ordinary kernel estimate or without a data taper fails.

Another application is a bootstrap procedure in the frequency domain. Figure 2 shows how the values of the periodogram fluctuate randomly around the (unknown) spectral density. One may use this fact to create more "periodogram samples" of the same kind by resampling in

a suitable manner from $I_T(\lambda) / f_T(\lambda)$ where $f_T(\lambda)$ is e.g. a kernel estimate. These additional samples may be for example used to get information on the distributional properties of estimates that are functionals of the periodogram. Franke and Härdle (1992) use such an approach to determine an adaptive bandwidth for a kernal estimate which has certain optimality properties.

Janas (1992) has constructed a wild bootstrap (cp. Mammen, 1992) which emulates also the fourth order moment structure of the estimates. He also proved that the classical bootstrap leads to a better approximation of the distribution of so called "ratio statistics" than the classical normal approximation does.

Although some of the procedures discussed above seem to be natural from our knowledge of the behaviour of iid-data, the weak dependence of the periodogram ordinates causes serious trouble in the exact proof of the results.

Acknowledgement. The simulations (apart from the prediction example in section 3) were done with the program SPECSIM (a program for the simulation of time series and spectral estimates). SPECSIM was originally created jointly with E. Ioannidis on the basis of the framework "Random & Template" by G. Sawitzki. Large contributions were made by M. Diller, D. Reinwald and R. v. Sachs. M. Brockmann contributed the local bandwidth selection procedure. I am grateful to all of them for their excellent work.

References

Brillinger, D. (1981). Time Series: Data Analysis and Theory. Mc Graw-Hill, Inc., New York.

Brockmann, M. (1992). Local Bandwidth Selection in Nonparametric Kernel Regression. PhD thesis, University of Heidelberg.

Brockwell, P. and Davis, R. (1991). Time Series: Theory and Methods. Springer, New York.

Burg, J. P. (1967). Maximum Entropy Spectral Analysis. Proceedings of the 37th Meeting of the Society of Exploration Geophysicists.

Dahlhaus, R. (1984). Parameter estimation of stationary processes with spectra containing strong peaks. Robust and Nonlinear Time Series Analysis. *Lecture Notes in Statist.* **26** Springer, New York 50 - 86.

Dahlhaus, R. (1988) Small sample effects in time series analysis: A new asymptotic theory and a new estimate. *Ann. Statist.*, Vol. **16**, No. 2 808 - 841.

Dahlhaus, R. (1990) Nonparametric High Resolution Spectral Estimation. Probab. *Th. Rel. Fields* **85** 147 - 180 .

Franke, J. and Härdle, W. (1992). On Bootstrapping Kernel Spectral Estimates. *Ann. Statist.* **20** 121 - 145.

Gasser,Th. , Kneip, A. and Köhler, W. (1991). A flexible and fast method for automatic smoothing. *J. Amer. Statist. Ass.* **86** 643 - 652.

Hainz, G. (1992). Asymptotische Eigenschaften des Burg-Schätzers. Diplomarbeit Universität Heidelberg.

Ioannidis, E. (1993). On the Asymptotic behaviour of the Capon Estimator. PhD thesis, University of Heidelberg.

Janas, D. (1992). Bootstrap Procedures for Time Series. Preprint SFB.

Mammen, E. (1992) When does bootstrap work: Asymptotic results and simulations. *Lecture Notes in Statist.* **77**, Springer Verlag.

Parzen, E. (1983). Autoregressive spectral estimation. In *Handbook of Statistics* (D. R. Brillinger and P. R. Krishnaiah, eds.) **3** North-Holland, Amsterdam 221- 247.

Priestley, M.B. (1981). Spectral Analysis and Time Series, vol. 1. Academic Press, London.

Tukey, J. W. (1967). An introduction to the calculations of numerical spectrum analysis. In: *Advanced Seminar on Spectral Analysis of Time Series* (B. Harris, ed.) Wiley, New York 25 - 46.

v. Sachs, R. (1991). Peak-Insensitive Nonparametric Spectrum Estimation. PhD thesis, University of Heidelberg.

Whittle, P. (1953). Estimation and information in stationary time series. *Ark. Mat.* **2** 423-434.